A Synopsis of the Thesis

DEVELOPMENT OF TEXT MINING ALGORITHUM FOR SENTIMENT ANALYSIS IN SOCIAL NETWORK USING NLP

Submitted to the

Swami Ramanand Teerth Marathwada University, Nanded

For the award of Degree of

Doctor of Philosophy in Computer Science

By

Mr.Patel Vasim Nizamoddin

Mobile: 8660371608

E-Mail: vasim.np@gmail.com

Category: Open

Under the Guidance of

Dr. B.G. Kodge

Mobile: 9923229672
E-mail: kodgebg@gmail.com
Swami Vivekananda College,Udgir

Research Centre:
Swami Vivekananda College,Udgir

DEVELOPMENT OF TEXT MINING ALGORITHUM FOR SENTIMENT ANALYSIS IN SOCIAL NETWORK USING NLP

1. INTRODUCTION

There are 50 to 60% of people worldwide use Facebook, whatsapp, YouTube, Instagram,twitter.There is huge amount of data generated at every moment and it is getting extremely difficult to get the relevant insights out of all that clutter. Word embedding tool is used to analyze online conversations and determine deeper context - positive, negative, neutral. Word embeddings are representations of words as vectors, learned by exploiting vast amounts of text. Each word is mapped to one vector and the vector values are learned in a way that resembles an artificial neural network.

The word embedding algorithm takes as its input from a large corpus of text and produces these vector spaces, typically of several hundred dimensions. A neural language model is trained on a large and the output of the network is used to each unique word to be assigned to a corresponding vector.

2. REVIEW OF LITERATURE

A recent popular news on NLP is the controversy that OpenAI has published a new GPT-2 language model but they refused to open source the full model due to its potential dark uses! It was trained via 8 million web pages and GPT-2 can *generate* long paragraphs of human-like coherent text and has potential to create fake news or spoof online identities. It was basically found too dangerous to make public.

Automatic analysis in on-line text Erik Boiy; Pieter Hens; Koen Deschacht; Marie-Francine Moens.The idea of automatic text analysis is important for marketing research, where companies wish to find out what the world thinks of their product; for monitoring newsgroups and forums, where fast and automatic detection of flaming is necessary; for analysis of customer feedback; or as informative augmentation for search engines.

Recently there was also news that Google has open sourced its natural language processing (NLP) pre-training model called bidirectional encoder representations from transformers (BERT). Then Baidu announced its own pre-trained NLP model called “ERNIE”.

The large tech companies and publishers including Facebook or Google Jigsaw are trying to find ways to detoxify the abundant abuse and harassment on the Internet. Though thousands of human moderators are still needed to avoid scandals until AI and NLP catch up.

3. SCOPE OF WORK

The proposed research can reach out to unhappy customers or help run a marketing/political campaign. Maybe you want to avoid crises or identify the top influencers

Writers word Analysis can help craft all this exponentially growing unstructured text into structured data using word embedding tools

These pre-trained NLP capabilities could be easily reapplied with emails, Facebook, Twitter, Instagram,YouTube, IMDB,eRetailer reviews, news

By analyzing sentiments from different sources one can predict the outcome of the event. Eg: in Election. It enables managers to track how voters feel about different issues, how they relate to speeches and actions of the candidate

4. OBJECTIVES OF RESEARCH

- Identifying financials risks or fraud
- proposed research can analyze text snippets in seconds.
- This research process natural language in order to perform various human like tasks like language translation or answering questions.
- Matching online advertisements, sentiment analysis for marketing or finance/trading
- Pre-trained and deep learning powered algorithms, act like humans and can be deployed much faster using reasonable computing resources
- Coherent text and has potential to create fake news or spoof online identities. It was basically found too dangerous to make public.
- Monitoring reputation of a specific brand. Eg: Twitter, Facebook.

. 5. METHODOLOGY TO BE ADOPTED:

Natural Language Processing (NLP) is a field at the intersection of computer science, artificial intelligence, and linguistics. The goal is for computers to process or "understand" natural language in order to perform various human like tasks like language translation or answering questions

The genius behind NLP is a concept called word embedding. Word embeddings are representations of words as vectors, learned by exploiting vast amounts of text. Each word is mapped to one vector and the vector values are learned in a way that resembles an artificial neural network One of the readily available pre-trained algorithms is called VADER (Valence Aware Dictionary and sentiment Reasoner) that is a lexicon(dictionary of sentiments in this case) and a simple rule-based model for general sentiment analysis. Its algorithms are optimized to sentiments expressed in social media like Twitter, online news, movie/product reviews etc. VADER can give us a Positivity and Negativity score that can be standardized in a range of -1 to 1.

6. REFERENCES:

[1] Ahmad Kamal , ―Subjectivity Classification using Machine Learning Techniques for Mining Feature - Opinion Pairs from Web Opinion Sources‖, International Journal of Computer Science Issues (IJCSI), Volume 10 Issue 5, 2013, pp 191-200.

[2] Aashutosh Bhatt, Ankit Patel, Harsh Chheda, Kiran Gawande, ―Amazon Review Classification and Sentiment Analysis ―,IJCSIT, Vol. 6 (6) , 2015, 5107-5110

[3] Ms.S.Yamini, Dr.V.Khanaa and Dr.Krishna Mohanta published a paper titled ―A state of the art review on various data mining techniques‖ in International Journal of Innovative Research in Science, Engineering and Technology (IJIESET), Vol.5, Issue 3, March 2016.Impact factor- 5.482.

[4] Liu B (2014) ,‖The science of detecting fake reviews‖, http://content26.com/blog/bing-liu-the-science-of-detecting-fakereviews.

[5]Baccianella,S.; Esuli, A.; Sebastiani, F. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Proceeding of the International Conference on Languguage Resources and Evaluation, Valletta. Malta, 17-23 May 2010; pp. 2200-2204. 49.

[6] Khan, F.H.; Bashir, S.; Qamar, U. Tom: Twitter opinion miming framework using hybrid classification scheme. Decis. Support Syst. 2014, 57, 245-257.

[7] Khatua, A.; Khatua, A,; Ghosh, k,; Chaki,N. Can# twitter_ trends predict election results? Evidence from 2014 indian general election. In Proceeding of the 48th Hawaii International Conference on System Sciences, Kauai, HI, USA, 5-8 January 2015; pp. 1676-1685.51.

[8] Hansen, L.K.; Arvidsson, A.; Nielsen, F.A.; Colleoni, E.;Etter, M. Good friends, bad news- affect and virality in twitter. In Future Information Technology; Springer: Berlin/Heidelberg, Germany, 2011;pp.34-43.52.hu, M.;

[9] Liu, B. Mining and summarizing customer reviews. In Proceedings of the Tenth ACMSIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA,USA, 22-25 August 2004; pp. 168-177.53.

[10] Saha, S.; Yadav, J.; Ranjan, P. Proposed approach for sarcasm detection in twitter. Indian. J. Sci. Technol.2017,10

Signature	Signature
Mr. Patel Vasim.N	Dr.Kodge B.G
(Research Scholar)	(Research Guide)